Preference-Based Rank Elicitation using Statistical Models: The Case of Mallows

Róbert Busa-Fekete, Eyke Hüllermeier, Balázs Szörényi

Preference-Based Rank Elicitation using Statistical Models: The Case of Mallows

Róbert Busa-Fekete[1] BUSAROBI@INF.U-SZEGED.HU
Eyke Hüllermeier[2] EYKE@UPB.DE
Balázs Szörényi[1,3] SZORENYI@INF.U-SZEGED.HU
[1]MTA-SZTE Research Group on Artificial Intelligence, Tisza Lajos krt. 103., H-6720 Szeged, Hungary
[2]Department of Computer Science, University of Paderborn, Warburger Str. 100, 33098 Paderborn, Germany
[3]INRIA Lille - Nord Europe, SequeL project, 40 avenue Halley, 59650 Villeneuve d'Ascq, France

Abstract

We address the problem of rank elicitation assuming that the underlying data generating process is characterized by a probability distribution on the set of all rankings (total orders) of a given set of items. Instead of asking for complete rankings, however, our learner is only allowed to query pairwise preferences. Using information of that kind, the goal of the learner is to reliably predict properties of the distribution, such as the most probable top-item, the most probable ranking, or the distribution itself. More specifically, learning is done in an online manner, and the goal is to minimize sample complexity while guaranteeing a certain level of confidence.

1. Introduction

Exploiting revealed preferences to learn a ranking over a set of options is a challenging problem with many practical applications. For example, think of crowd-sourcing services like the Amazon Mechanical Turk, where simple questions such as pairwise comparisons between decision alternatives are asked to a group of annotators. The task is to approximate an underlying target ranking on the basis of these pairwise comparisons, which are possibly noisy and partially inconsistent (Chen et al., 2013). Another application worth mentioning is the ranking of XBox gamers based on their pairwise online duels; the ranking system of XBox is called TrueSkill™(Guo et al., 2012).

In this paper, we focus on a problem that we call *preference-based rank elicitation*. In the setting of this problem, we proceed from a finite set of items $\mathcal{I} = \{1, \ldots, M\}$ and assume a fixed but unknown probability distribution $\mathbb{P}(\cdot)$ to be defined on the set of all rankings (total orders) $\mathbf{r}$ of these items; for example, one may think of $\mathbb{P}(\mathbf{r})$ as the probability that an individual, who is randomly chosen from a population, reports the preference order $\mathbf{r}$ over the items $\mathcal{I}$. However, instead of asking for full rankings, we are only allowed to ask for the comparison of pairs of items. The goal, then, is to quickly gather enough information so as to enable the reliable prediction of properties of the distribution $\mathbb{P}(\cdot)$, such as the most probable top-item, the most probable ranking, or the distribution itself. More specifically, learning is done in an online manner, and the goal is to minimize sample complexity while guaranteeing a certain level of confidence.

After a brief survey of related work, we introduce notation in Section 3 and describe our setting more formally in Section 4. In Section 5, we recall the well-known Mallows ϕ-model, which is the model we assume for the distribution $\mathbb{P}(\cdot)$ in this paper. In Section 6, we introduce and analyze rank elicitation algorithms for the problems mentioned above. In Section 7, we present an experimental study, and finally conclude the paper in Section 8.

2. Related work

Pure exploration algorithms for the stochastic multi-armed bandit problem sample the arms a certain number of times (not necessarily known in advance), and then output a recommendation, such as the best arm or the m best arms (Bubeck et al., 2009; Even-Dar et al., 2002; Bubeck et al., 2013; Gabillon et al., 2011; Cappé et al., 2012). While our algorithm can be seen as a pure exploration strategy, too, we do not assume that *numerical* feedback can be generated for *individual* options; instead, our feedback is *qualitative* and refers to *pairs* of options.

Different types of preference-based multi-armed bandit setups have been studied in a number of recent publications. Like in our case, the (online) learner compares arms in a pairwise manner, and the (stochastic) outcome of a com-

Proceedings of the 31^{st} International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

parison essentially informs about whether or not an option is preferred to an other one. We can classify these works into two main groups. Approaches from first group, such as (Yue et al., 2012) and (Yue & Joachims, 2011), assume certain regularity properties for the pairwise comparisons, such as strong stochastic transitivity, thereby assuring the existence of a natural target ranking. The second group does not make such assumptions, and instead derives a target ranking from the pairwise relation by means of a ranking rule; for example, (Busa-Fekete et al., 2013) and (Urvoy et al., 2013) are of that kind. Our work is obviously closer to the first group, since we assume that preferences are generated by the Mallows model (Mallows, 1957)—as will be seen later on, this assumption implies specific regularity properties on the pairwise comparisons, too.

There is a vast array of papers that devise algorithms related to the Mallows ϕ-model. Our work is specifically related to Lu & Boutilier (2011), who aim at learning the Mallows model based on pairwise preferences. Their technique allows for sampling the posterior probabilities of the Mallows model conditioned on a set of pairwise observations. In this paper, however, we consider the online setting, where the learner needs to decide which pairs of options to compare next.

Braverman & Mossel (2008) solve the Kemeny (rank aggregation) problem when the distribution of rankings belongs to the family of Mallows. The authors prove that, in this special case, the problem is less complex than in the general case and can be solved in polynomial time.

Jamieson & Nowak (2011) consider an online learning setup with the goal to learn an underlying ranking via sampling of noisy pairwise preferences. However, they assume that the objects to be ranked can be embedded in a d-dimensional Euclidean space, and that the rankings reflect their relative distances from a common reference point in $\mathbb{R}^d$. The authors introduce an adaptive sampling algorithm, which has an expected sample complexity of order $d \log n$.

3. Notation

A set of options/objects/items to be ranked is denoted by $\mathcal{I}$. To keep the presentation simple, we assume that items are identified by natural numbers, so $\mathcal{I} = [M] = \{1, \ldots, M\}$. A *ranking* is a bijection $\mathbf{r}$ on $\mathcal{I}$, which can also be represented as a vector $\mathbf{r} = (r_1, \ldots, r_M) = (\mathbf{r}(1), \ldots, \mathbf{r}(M))$, where $r_j = \mathbf{r}(j)$ is the rank of the jth item. The set of rankings can be identified with the symmetric group $\mathbb{S}_M$ of order M. Each ranking $\mathbf{r}$ naturally defines an associated *ordering* $\mathbf{o} = (o_1, \ldots, o_M) \in \mathbb{S}_M$ of the items, namely the inverse $\mathbf{o} = \mathbf{r}^{-1}$ defined by $o_{\mathbf{r}(j)} = j$ for all $j \in [M]$.

For a permutation $\mathbf{r}$, we write $\mathbf{r}(i, j)$ for the permutation in which r_i and r_j, the ranks of items i and j, are replaced with each other. We denote by $\mathcal{L}(r_i = j) = \{\mathbf{r} \in \mathbb{S}_M \,|\, r_i = j\}$ the subset of permutations for which the rank of item i is j, and by $\mathcal{L}(r_j > r_i) = \{\mathbf{r} \in \mathbb{S}_M \,|\, r_j > r_i\}$ those for which the rank of j is higher than the rank of i, that is, item i is preferred to j, written $i \succ j$.

We assume $\mathbb{S}_M$ to be equipped with a probability distribution $\mathbb{P} : \mathbb{S}_M \to [0, 1]$; thus, for each ranking $\mathbf{r}$, we denote by $\mathbb{P}(\mathbf{r})$ the probability to observe this ranking. Moreover, for each pair of items i and j, we denote by

$$p_{i,j} = \mathbb{P}(i \succ j) = \sum_{\mathbf{r} \in \mathcal{L}(r_j > r_i)} \mathbb{P}(\mathbf{r}) \tag{1}$$

the probability that i is preferred to j (in a ranking randomly drawn according to $\mathbb{P}$). We denote the matrix of $p_{i,j}$ values by $\mathbf{P} = [p_{i,j}]_{1 \le i,j \le M}$.

4. Preference-based rank elicitation

Our learning problem consists of making a good prediction about $\mathbb{P}$. Concretely, we consider three different goals of the learner, depending on whether the application calls for the prediction of a single item, a full ranking of items or the entire probability distribution:

MPI: Find the most preferred item i^*, namely the item whose probability of being top-ranked is maximal:

$$\begin{aligned} i^* &= \operatorname*{argmax}_{1 \le i \le M} \mathbb{E}_{\mathbf{r} \sim \mathbb{P}} [\![r_i = 1]\!] \\ &= \operatorname*{argmax}_{1 \le i \le M} \sum_{\mathbf{r} \in \mathcal{L}(r_i = 1)} \mathbb{P}(\mathbf{r}) \end{aligned}$$

where $[\![\cdot]\!]$ is the indicator function which is 1 if its argument is true and 0 otherwise.

MPR: Find the most probable ranking $\mathbf{r}^*$:

$$\mathbf{r}^* = \operatorname*{argmax}_{\mathbf{r} \in \mathbb{S}_M} \mathbb{P}(\mathbf{r})$$

KLD: Produce a good estimate $\widehat{\mathbb{P}}$ of the distribution $\mathbb{P}$, that is, an estimate with small KL divergence:

$$\mathrm{KL}\left(\mathbb{P}, \widehat{\mathbb{P}}\right) < \epsilon$$

All three goals are meant to be achieved with probability at least $1 - \delta$. Our learner operates in an online setting. In each iteration, it is allowed to gather information by asking for a single *pairwise comparison* between two items. Thus, it selects two items i and j, and then observes either preference $i \succ j$ or $j \succ i$; the former occurs with probability $p_{i,j}$ as defined in (1), the latter with probability $p_{j,i} = 1 - p_{i,j}$. Based on this observation, the learner updates its estimates and decides either to continue the learning process or to

terminate and return its prediction. What we are mainly interested in is the sample complexity of the learner, that is, the number of pairwise comparisons it queries prior to termination.

5. Mallows ϕ-model

So far, we did not make any assumptions about the probability distribution $\mathbb{P}$ on $\mathbb{S}_M$. Without any restriction, however, efficient learning is arguably impossible. Subsequently, we shall therefore assume that $\mathbb{P}$ is a Mallows model (Mallows, 1957), one of the most well-known and widely used statistical models of rank data (Marden, 1995). The Mallows model or, more specifically, Mallow's ϕ-distribution is a parameterized, distance-based probability distribution that belongs to the family of exponential distributions:

$$\mathbb{P}(\mathbf{r} \,|\, \theta, \widetilde{\mathbf{r}}) = \frac{1}{Z(\phi)}\, \phi^{d(\mathbf{r},\widetilde{\mathbf{r}})} \tag{2}$$

where ϕ and $\widetilde{\mathbf{r}}$ are the parameters of the model: $\widetilde{\mathbf{r}} = (\tilde{r}_1, \ldots, \tilde{r}_M) \in \mathbb{S}_M$ is the location parameter (center ranking) and $\phi \in (0, 1]$ the spread parameter. Moreover, $d(\cdot, \cdot)$ is the Kendall distance on rankings, that is, the number of discordant item pairs:

$$d(\mathbf{r}, \widetilde{\mathbf{r}}) = \sum_{1 \le i < j \le M} [\![(r_i - r_j)(\tilde{r}_i - \tilde{r}_j) < 0]\!] \ .$$

The normalization factor in (2) can be written as

$$Z(\phi) = \sum_{\mathbf{r} \in \mathbb{S}_M} \mathbb{P}(\mathbf{r} \,|\, \theta, \widetilde{\mathbf{r}}) = \prod_{i=1}^{M-1} \sum_{j=0}^{i} \phi^j$$

and thus only depends on the spread (Fligner & Verducci, 1986). Note that, since $d(\mathbf{r}, \widetilde{\mathbf{r}}) = 0$ is equivalent to $\mathbf{r} = \widetilde{\mathbf{r}}$, the center ranking $\widetilde{\mathbf{r}}$ is the mode of $\mathbb{P}(\cdot \,|\, \theta, \widetilde{\mathbf{r}})$, that is, the most probable ranking according to the Mallows model.

6. Algorithms

Before tackling the problems introduced above (MPI, MPR, KLD), we need some additional notation. The pair of items chosen by the learner in iteration t is denoted (i^t, j^t), and the feedback received is defined as $o^t = 1$ if $i^t \succ j^t$ and $o^t = 0$ if $j^t \succ i^t$. The set of steps among the first t iterations in which the learner decides to compare items i and j is denoted by $I^t_{i,j} = \{\ell \in [t] \,|\, (i^\ell, j^\ell) = (i, j)\}$, and the size of this set by $n^t_{i,j} = \# I^t_{i,j}$.[1] The proportion of "wins" of item i against item j up to iteration t is then given by

$$\widehat{p}^t_{i,j} = \frac{1}{n^t_{i,j}} \sum_{\ell \in I^t_{i,j}} o^\ell \ .$$

[1] We omit the index t if there is no danger of confusion.

Since our samples are i.i.d., $\widehat{p}^t_{i,j}$ is an estimate of the pairwise probability (1).

6.1. The most preferred item (MPI)

We start with a simple observation on the Mallows ϕ-model regarding item i^*, which is ranked first with the highest probability.

Proposition 1. *For a Mallows ϕ-model with parameters ϕ and $\widetilde{\mathbf{r}}$, it holds that $\widetilde{r}_{i^*} = 1$.*

Proof. Let $\widetilde{r}_i = 1$ for some i, and consider the following difference for some $j \neq i$:

$$\begin{aligned} \sum_{\mathbf{r} \in \mathcal{L}(r_i = 1)} \mathbb{P}(\mathbf{r} \,|\, \phi, \widetilde{\mathbf{r}}) - \sum_{\mathbf{r} \in \mathcal{L}(r_j = 1)} \mathbb{P}(\mathbf{r} \,|\, \phi, \widetilde{\mathbf{r}}) &= \\ = \sum_{\mathbf{r} \in \mathcal{L}(r_i = 1)} \mathbb{P}(\mathbf{r} \,|\, \phi, \widetilde{\mathbf{r}}) - \mathbb{P}(\mathbf{r}(i,j) \,|\, \phi, \widetilde{\mathbf{r}}) & \\ = \frac{1}{Z(\phi)} \sum_{\mathbf{r} \in \mathcal{L}(r_i = 1)} \phi^{d(\mathbf{r},\widetilde{\mathbf{r}})} - \phi^{d(\mathbf{r}(i,j),\widetilde{\mathbf{r}})} \ , & \end{aligned}$$

which is always always bigger than zero, if $d(\mathbf{r}, \widetilde{\mathbf{r}}) < d(\mathbf{r}(i,j), \widetilde{\mathbf{r}})$ for all $\mathbf{r} \in \mathcal{L}(r_i = 1)$. To show that $d(\mathbf{r}, \widetilde{\mathbf{r}}) < d(\mathbf{r}(i,j), \widetilde{\mathbf{r}})$ for a $\mathbf{r} \in \mathcal{L}(r_i = 1)$ is very technical, thus the proof of this claim is deferred to the supplementary material (see Appendix A). This completes the proof. □

Next, we recall a result of Mallows (1957), stating that the matrix $\mathbf{P}$ has a special form for a Mallows ϕ-model: permutating its rows and columns based on the center ranking, it is Toeplitz, and its entries can be calculated analytically as functions of the model parameters ϕ and $\widetilde{\mathbf{r}}$.

Theorem 2. *Assume the Mallows model with parameters ϕ and $\widetilde{\mathbf{r}}$. Then, for any pair of items i and j such that $\widetilde{r}_i < \widetilde{r}_j$, the marginal probability (1) is given by $p_{i,j} = g(\widetilde{r}_i, \widetilde{r}_j, \phi)$, where*

$$g(i, j, \phi) = h(j - i + 1, \phi) - h(j - i, \phi)$$

with $h(k, \phi) = k/(1 - \phi^k)$.

The following corollary summarizes some consequences of Theorem 2 that we shall exploit in our implementation.

Corollary 3. *For a given Mallows ϕ-model with parameters ϕ and $\widetilde{\mathbf{r}}$, the following claims hold:*

1. *For any pair of items $i, j \in [M]$ such that $\widetilde{r}_i < \widetilde{r}_j$, the pairwise marginal probabilities satisfy $p_{i,j} \geq \frac{1}{1+\phi} > 1/2$ with equality holding iff $\widetilde{r}_i = \widetilde{r}_j - 1$. Moreover, for items i, j, k satisfying $\widetilde{r}_i = \widetilde{r}_j - \ell = \widetilde{r}_k - \ell - 1$ with $1 < \ell$, it holds that $p_{i,j} - p_{i,k} = \mathcal{O}(\ell \phi^\ell)$.*
2. *For any pair of items $i, j \in [M]$ such that $\widetilde{r}_i \leq \widetilde{r}_j + 1$ the pairwise marginal probabilities satisfy $p_{i,j} \leq$*

$\frac{\phi}{1+\phi} < 1/2$ *with equality holding iff* $\widetilde{r}_i = \widetilde{r}_j + 1$. *Moreover, for items* i, j, k *satisfying* $\widetilde{r}_i = \widetilde{r}_j + \ell = \widetilde{r}_k + \ell + 1$ *with* $1 < \ell$, *it holds that* $p_{i,k} - p_{i,j} = \mathcal{O}(\ell\phi^\ell)$.

3. *For any* $i, j \in [M]$ *such that* $i \neq j$, $p_{i,j} > 1/2$ *iff* $\widetilde{r}_i < \widetilde{r}_j$, *and* $p_{i,j} < 1/2$ *iff* $\widetilde{r}_i > \widetilde{r}_j$. *Therefore for any item* $i \in [M]$, $\#A_{i+} = \widetilde{r}_i - 1$, *and* $\#A_{i-} = M - \widetilde{r}_i$ *where* $A_{i+} = \{j \in [M] | p_{i,j} > 1/2\}$ *and* $A_{i-} = \{j \in [M] | p_{i,j} < 1/2\}$.

Proof. To show the first claim, consider a pair of items $i, j \in [M]$ for which $\widetilde{r}_i = \widetilde{r}_j - 1$. Then, based on Theorem 2, a simple calculation yields $p_{i,j} = g(\widetilde{r}_i, \widetilde{r}_j, \phi) = h(2, \phi) - h(1, \phi) = \frac{1}{1+\phi}$. It is also easy to show that $h(\cdot, \phi)$ is a strictly increasing convex function for any $\phi \in (0, 1]$. This can be checked by showing first that $h(x) = x/(1 - e^x)$ is a strictly increasing convex function, and then by applying the transformation[2] $x/(1 - \phi^x) = h(x \log(1/\phi))/\log(1/\phi)$. And thus $h(\ell + 2, \phi) - h(\ell + 1, \phi) > h(\ell + 1, \phi) - h(\ell, \phi)$ for any $\ell > 0$. From this, using induction, one obtains that $p_{i,k} > p_{i,j}$ whenever $\widetilde{r}_k > \widetilde{r}_j > \widetilde{r}_i$. To complete the proof for the first claim define $f(x) = x - x/(1 + \phi^x) = x\phi^x/(1 + \phi)$, and note that for indices i, j, k satisfying the requirements of the claim it holds that $p_{i,j} - p_{i,k} = f(\ell + 2) + f(\ell) - 2f(\ell + 1)$.

The proof of the second claim is analogous to the first one, noting that $p_{i,j} = 1 - p_{j,i}$ for all $i, j \in [M]$. The third claim is a consequence of the first two claims. □

Based on Theorem 2 and Corollary 3, one can devise an efficient algorithm for identifying the most preferred item when the underlying distribution is Mallows. The pseudo-code of this algorithm, called MALLOWSMPI, is shown in Algorithm 1. It maintains a set of active indices A, which is initialized with all items $[M]$. In each iteration, it picks an item $j \in A$ at random and compares item i to j until the confidence interval of $\widehat{p}_{i,j}$ does not contain $1/2$. Finally, it keeps the winner of this pairwise duel (namely item i if $\widehat{p}_{i,j}$ is significantly bigger than $1/2$ and item j otherwise).[3] This simple strategy is suggested by Corollary 3, which shows that the "margin" $\min_{i \neq j} |1/2 - p_{i,j}|$ around $1/2$ is relatively wide; more specifically, there is no $p_{i,j} \in (\frac{\phi}{1+\phi}, \frac{1}{1+\phi})$. Moreover, deciding whether an item j has higher or lower rank than i (with respect to $\widetilde{\mathbf{r}}$) is easier than selecting the preferred option from two candidates j and k for which $j, k \neq i$ (see Corollary 3).

As an illustration, Figure 1 shows a plot of the matrix $\mathbf{P}$ for a Mallows ϕ-model. As can be seen, the surface is steepest close to the diagonal, which is in agreement with our above remarks about the "margin".

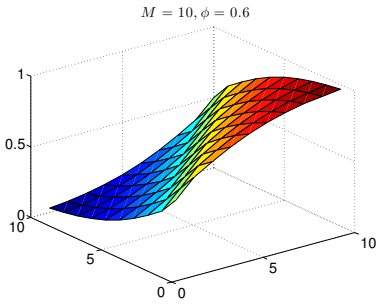


Figure 1. The pairwise marginal probability matrix $\mathbf{P}$ for a Mallows ϕ-model (with $\widetilde{\mathbf{r}}$ the identity, $\phi = 0.6$, $M = 10$) calculated based on Theorem 2.

Algorithm 1 MALLOWSMPI(δ)

```
Set A = {1, ..., M}
Pick a random index i ∈ A and set A = A \ {i}
while A ≠ ∅ do
    Pick a random index j ∈ A and set A = A \ {j}
    repeat
        Observe o = [[ r_i < r_j ]]
        p̂_{i,j} = p̂_{i,j} + o, n_{i,j} = n_{i,j} + 1
        c_{i,j} = sqrt( 1/(2 n_{i,j}) log 4M n_{i,j}^2 / δ )
    until 1/2 ∉ [p̂_{i,j} − c_{i,j}, p̂_{i,j} + c_{i,j}]
    if 1/2 > p̂_{i,j} + c_{i,j} then        ▷ r̃_j < r̃_i w.h.p.
        i = j
return i
```

Similarly to the sample complexity analysis given by Even-Dar et al. (2002) for PAC-bandits, we can upper-bound the number of pairwise comparisons taken by MALLOWSMPI with high probability.

Theorem 4. *Assume the Mallows model with parameters* ϕ *and* $\widetilde{\mathbf{r}}$ *as an underlying ranking distribution. Then, for any* $0 < \delta < 1$, MALLOWSMPI *outputs the most preferred item with probability at least* $1 - \delta$, *and the number of pairwise comparison taken is*

$$\mathcal{O}\left(\frac{M}{\rho^2} \log \frac{M}{\delta\rho}\right) ,$$

where $\rho = \frac{1-\phi}{1+\phi}$.

Proof. First note that by setting the length of the confidence interval to $c_{i,j} = \sqrt{1/2n_{i,j} \log(4Mn_{i,j}^2/\delta)}$, we have

$$\mathbb{P}\left(|p_{i,j} - \widehat{p}_{i,j}| \geq c_{i,j}\right) \leq 2\exp(-2c_{i,j}^2 n_{i,j}) = \frac{\delta}{2Mn_{i,j}^2}$$

[2] Throughout the paper, $\log(x)$ denotes a natural logarithm.

[3] In contrast to the INTERLEAVED FILTER (Yue et al., 2012), which compares all active options to each other, we only compare two options at a time.

for any time step. Therefore, $p_{i,j} \in [\widehat{p}_{i,j} - c_{i,j}, \widehat{p}_{i,j} + c_{i,j}]$ for any pair of items in every time step with probability at least $1 - \delta/M$. Moreover, according to Corollary 3, if $p_{i,j} > 1/2$, then $\widetilde{r}_i < \widetilde{r}_j$, and $p_{i,j} < 1/2$ implies $\widetilde{r}_i > \widetilde{r}_j$, therefore we always keep the item which has lower rank with respect to $\widetilde{r}$ with probability at least $1 - \delta/M$. In addition, since at most $M - 1$ distinct pairs of items are compared (always retaining the more preferred one), the algorithm outputs the most preferred item with probability at least $1 - \delta$.

To calculate the sample complexity, based on Corollary 3, we know that $p_{i,j} \notin (\frac{\phi}{1+\phi}, \frac{1}{1+\phi})$. Therefore to achieve that $1/2 \notin [\widehat{p}_{i,j} - c_{i,j}, \widehat{p}_{i,j} + c_{i,j}]$ where $p_{i,j} > 1/2$, the following has to be satisfied:

$$\sqrt{\frac{1}{2n_{i,j}} \log \frac{4Mn_{i,j}^2}{\delta}} < \left(\frac{1}{1+\phi} - \frac{1}{2}\right) = \frac{1-\phi}{2(1+\phi)}$$

To achieve this, simple calculation yields that the number of samples that is needed, is

$$\left\lceil \frac{4}{\rho^2} \log \frac{4M}{\delta} + \frac{4}{\rho^2} \left(1 + 2 \log \frac{4}{\rho^2}\right) \right\rceil = \mathcal{O}\left(\frac{1}{\rho^2} \log \frac{M}{\delta\rho}\right)$$

if $p_{i,j} \in [\widehat{p}_{i,j} - c_{i,j}, \widehat{p}_{i,j} + c_{i,j}]$. A similar argument applies in the case $p_{i,j} < 1/2$, which completes the proof. □

6.2. The most probable ranking (MPR)

For a Mallows ϕ-model, the center ranking coincides with the mode of the distribution. Moreover, based on Corollary 3, we know that $p_{i,j} > 1/2$ if (and only if) an item i precedes an item j in the center ranking $\widetilde{\mathbf{r}}$. Therefore, finding the most probable ranking amounts to solving a sorting problem in which the order of two items needs to be decided with high probability. The implementation of our method is shown in Algorithm 2, which is based on the well-known merge sort algorithm. Accordingly, it calls a recursive procedure MMREC, given in Procedure 3, which divides the unsorted set of items into two subsets, calls itself recursively, and finally merges the two sorted list returned by calling the procedure MALLOWSMERGE shown in Algorithm 4. The MALLOWSMERGE procedure merges the sorted item lists, and whenever the order of two items i and j is needed, it compares these items until the confidence interval for $p_{i,j}$ no longer overlaps $1/2$.

Algorithm 2 MALLOWSMPR(δ)

for $i = 1 \to M$ **do** $r_i = i, r'_i = 0$
$(\mathbf{r}', \mathbf{r}) = \text{MMREC}(\mathbf{r}, \mathbf{r}', \delta, 1, M)$
for $i = 1 \to M$ **do** $r_{r'_i} = i$
return $\mathbf{r}$

One can upper-bound the sample complexity of MALLOWSMPR in a similar way as for MALLOWSMPI.

Procedure 3 MMREC($\mathbf{r}, \mathbf{r}', \delta, i, j$)

if $j - i > 0$ **then**
  $k = \lceil (i + j)/2 \rceil$
  $(\mathbf{r}, \mathbf{r}') = \text{MMREC}(\mathbf{r}, \mathbf{r}', \delta, i, k - 1)$
  $(\mathbf{r}, \mathbf{r}') = \text{MMREC}(\mathbf{r}, \mathbf{r}', \delta, k, j)$
  $(\mathbf{r}, \mathbf{r}') = \text{MALLOWSMERGE}(\mathbf{r}, \mathbf{r}', \delta, i, k, j)$
  for $\ell = i \to j$ **do** $r_\ell = r'_\ell$
return $(\mathbf{r}, \mathbf{r}')$

Procedure 4 MALLOWSMERGE($\mathbf{r}, \mathbf{r}', \delta, i, k, j$)

$\ell = i, \ell' = k$
for $q = i \to j$ **do**
  if $(\ell < k) \& (\ell' \leq j)$ **then**
    repeat
      Observe $o = \mathbb{I}\{r_\ell < r_{\ell'}\}$
      $\widehat{p}_{\ell,\ell'} = \widehat{p}_{\ell,\ell'} + o, n_{\ell,\ell'} = n_{\ell,\ell'} + 1$
      $c_{\ell,\ell'} = \sqrt{\frac{1}{2n_{\ell,\ell'}} \log \frac{4n_{\ell,\ell'}^2 C_M}{\delta}}$
      with $C_M = \lceil M \log_2 M - 0.91392 \cdot M + 1 \rceil$
    until $1/2 \notin [\widehat{p}_{\ell,\ell'} - c_{\ell,\ell'}, \widehat{p}_{\ell,\ell'} + c_{\ell,\ell'}]$
    if $1/2 < \widehat{p}_{\ell,\ell'} - c_{\ell,\ell'}$ **then**
      $r'_q = r_\ell, \ell = \ell + 1$
    else
      $r'_q = r_{\ell'}, \ell' = \ell' + 1$
  else
    if $(\ell < k)$ **then**
      $r'_q = r_\ell, \ell = \ell + 1$
    else
      $r'_q = r_{\ell'}, \ell' = \ell' + 1$
return $(\mathbf{r}, \mathbf{r}')$

Theorem 5. *Assume the Mallows model with parameters ϕ and $\widetilde{\mathbf{r}}$ as an underlying ranking distribution. Then, for any $0 < \delta < 1$,* MALLOWSMPR *outputs the most probable ranking with probability at least $1 - \delta$, and the number of pairwise comparison taken by the algorithm is*

$$\mathcal{O}\left(\frac{M \log_2 M}{\rho^2} \log \frac{M \log_2 M}{\delta\rho}\right)$$

where $\rho = \frac{1-\phi}{1+\phi}$.

Proof. We adapted the two-way top-down merge sort algorithm whose worst case performance is upper bounded by $C_M = \lceil M \log_2 M - 0.91392 \cdot M + 1 \rceil$ (Theorem 1, Flajolet & Golin (1994)). Analogously to the proof of Theorem 4, by setting the confidence interval $c_{i,j}$ to $\sqrt{1/2n_{i,j} \log(n_{i,j}^2 4C_M/\delta)}$, it holds that for any pairs of items i and j, $p_{i,j} \in [\widehat{p}_{i,j} - c_{i,j}, \widehat{p}_{i,j} + c_{i,j}]$ for every time step with probability at least $1 - \delta/C_M$. According to Corollary 3, $p_{i,j} > 1/2$ implies $\widetilde{r}_i < \widetilde{r}_j$, and $p_{i,j} < 1/2$ implies $\widetilde{r}_i > \widetilde{r}_j$, in addition, at most C_M distinct pairs of

items are compared at most, therefore the algorithm outputs the most probable ranking with probability at least $1-\delta$.

Analogously to the proof of Theorem 4, the number of pairwise comparisons required by the MALLOWSMPR procedure to assure $1/2 \notin [\widehat{p}_{i,j} - c_{i,j}, \widehat{p}_{i,j} + c_{i,j}]$ for a pair of items i and j is $\mathcal{O}\left(\frac{1}{\rho^2} \log \frac{M \log_2 M}{\delta \rho}\right)$. Moreover, the worst case performance of merge sort is $\mathcal{O}(M \log_2 M)$, which completes the proof. □

In principle, sorting algorithms other than merge sort could be applied, too. For example, we put the implementation of the popular quick sort algorithm, called MALLOWSQUICK, in the supplementary material (see Appendix B), although its worst case complexity is not as good as the one of merge sort ($\mathcal{O}(M^2)$ instead of $\mathcal{O}(M \log M)$). Provided knowledge about how much the distribution of the number of pairwise comparisons concentrates around its mean for fixed M, one could also make use of the expected performance of sorting algorithms to prove PAC sample complexity bounds (like Theorem 5). As far as we know, however, there is no concentration result for its average performance with a fixed M.[4] For MALLOWSQUICK, we can therefore only prove a sample complexity bound of $\mathcal{O}\left(\frac{M^2}{\rho^2} \log \frac{M^2}{\delta \rho}\right)$. In Appendix E.1, we empirically compared MALLOWSQUICK with MALLOWSMPR in terms of sample complexity.

Remark 6. *The leading factor of sample complexity of* MALLOWSMERGE *differs from the one of* MALLOWSMPI *by a log factor. This was to be expected, and simply reflects the difference in worst case complexity for finding the best element in an array and sorting an array by using merge sort algorithm.*

6.3. Kullback-Leibler divergence (KLD)

In order to produce a model estimation that is close to the true Mallows model in terms of KL divergence, the parameters ϕ and $\widetilde{\mathbf{r}}$ must be estimated with an appropriate precision and confidence. First, by using MALLOWSMPR (see Algorithm 2), the center ranking $\widetilde{\mathbf{r}}$ can be found with probability at least $1-\delta$. For the sake of simplicity, we subsequently assume that this has already been done (actually with a corrected δ, as will be explained later).

Based on Corollary 3, we know that $p_{i,j} = \frac{1}{1+\phi}$ for a pair of items i and j such that $\widetilde{r}_i = \widetilde{r}_j + 1$. Assume that we are given an estimate $\widehat{p}_{i,j}$ with a confidence interval $c_{i,j}$ such that $\widetilde{r}_i < M$. Then,

$$\widehat{p}_{i,j} - c_{i,j} \leq \frac{1}{1+\phi} \leq \widehat{p}_{i,j} + c_{i,j}$$

implies the following confidence interval for ϕ:

$$\underbrace{\frac{1}{\widehat{p}_{i,j} + c_{i,j}} - 1}_{=\phi_L} \leq \phi \leq \underbrace{\frac{1}{\widehat{p}_{i,j} - c_{i,j}} - 1}_{=\phi_U} \tag{3}$$

Next, we upper-bound the KL divergence between two Mallows distributions $\mathbb{P}(\cdot \,|\, \phi_2, \widetilde{\mathbf{r}})$ and $\mathbb{P}(\cdot \,|\, \phi_2, \widetilde{\mathbf{r}})$ sharing the same center ranking:

$$\begin{aligned}\mathrm{KL}(\mathbb{P}(\cdot\,|\phi_1, \widetilde{\mathbf{r}}), \mathbb{P}(\cdot\,|\phi_2, \widetilde{\mathbf{r}})) \leq \\ \leq \frac{M(M-1)}{2} \log \frac{\phi_1}{\phi_2} + \log \frac{Z(\phi_2)}{Z(\phi_1)}\end{aligned} \tag{4}$$

Since the derivation of this result is fairly technical, it is deferred to the supplementary material (see Appendix C).

Equipped with a confidence interval $[\phi_L, \phi_U]$ for ϕ according to (3), we can upper-bound $\mathrm{KL}(\mathbb{P}(\cdot \,|\, \phi, \widetilde{\mathbf{r}}), \mathbb{P}(\cdot \,|\, \widehat{\phi}, \widetilde{\mathbf{r}}))$ for any $\widehat{\phi} \in [\phi_L, \phi_U]$ thanks to (4). Thus, with high probability, we have

$$\begin{aligned}&\mathrm{KL}(\mathbb{P}(\cdot \,|\, \phi, \widetilde{\mathbf{r}}), \mathbb{P}(\cdot \,|\, \widehat{\phi}, \widetilde{\mathbf{r}})) \\ &\quad\leq \frac{M(M-1)}{2} \log \frac{\phi}{\widehat{\phi}} + \log \frac{Z(\widehat{\phi})}{Z(\phi)} \\ &\quad\leq \frac{M(M-1)}{2} \log \frac{\phi_U}{\phi_L} + \log \frac{Z(\phi_U)}{Z(\phi_L)}\,,\end{aligned} \tag{5}$$

because $Z(.)$ is a monotone function. Based on (5), we can empirically test whether the confidence bound for ϕ is tight enough, such that any value in $[\phi_L, \phi_U]$ will define a distribution that is close to the true one (for this, we have to be aware of the center ranking with probability at least $1 - \delta/2$).

Algorithm 5 MALLOWSKLD(δ, ϵ)

1: $\widehat{\mathbf{r}}$ =MALLOWSMPR$(\delta/2)$
2: Pick a random pair of indices i and j for which $\widehat{r}_i < M$ and $\widehat{r}_i = \widehat{r}_j + 1$
3: **repeat**
4: Observe $o = \mathbb{I}\{r_i < r_j\}$
5: $\widehat{p}_{i,j} = \widehat{p}_{i,j} + o$, $n_{i,j} = n_{i,j} + 1$
6: $c_{i,j} = \sqrt{\frac{1}{2n_{i,j}} \log \frac{8n_{i,j}^2}{\delta}}$
7: $\phi_L = \frac{1}{\widehat{p}_{i,j} + c_{i,j}} - 1$, $\phi_U = \frac{1}{\widehat{p}_{i,j} - c_{i,j}} - 1$
8: **until** $\frac{M(M-1)}{2} \log \frac{\phi_U}{\phi_L} + \log \frac{Z(\phi_U)}{Z(\phi_L)} < \epsilon$
9: **return** $\widehat{\mathbf{r}}$ and any $\widehat{\phi} \in [\phi_L, \phi_U]$

Our implementation is shown in Algorithm 5. In a first step, it identifies the center ranking using MALLOWSMPR with probability at least $1 - \delta/2$. Then, it gradually estimates ϕ and terminates if the stopping condition based on (5) is satisfied. The sample complexity of MALLOWSKLD can be analyzed in the same way as for MALLOWSMPI and MALLOWSMPR. Due to space limitations, the proof is deferred to the supplementary material (see Appendix D).

[4] Although results on rates of convergence for the distribution of pairwise comparisons when $M \to \infty$ are available (Fill & Janson, 2002).

Theorem 7. *Assume that the ranking distribution is Mallows with parameters ϕ and $\widetilde{\mathbf{r}}$. Then, for any $\epsilon > 0$ and $0 < \delta < 1$,* MALLOWSKLD *returns parameter estimates $\widehat{\mathbf{r}}$ and $\widehat{\phi}$ for which $\mathrm{KL}(\mathbb{P}(\cdot \mid \phi, \widetilde{\mathbf{r}}), \mathbb{P}(\cdot \mid \widehat{\phi}, \widehat{\mathbf{r}})) < \epsilon$, and the number of pairwise comparisons requested by the algorithm is*

$$\mathcal{O}\left(\frac{M \log_2 M}{\rho^2} \log \frac{M \log_2 M}{\delta \rho} + \frac{1}{D(\epsilon)^2} \log \frac{1}{\delta D(\epsilon)}\right) ,$$

where $\rho = \frac{1-\phi}{1+\phi}$ and

$$D(\epsilon) = \frac{\phi}{6(\phi+1)^2}\left(1 - \frac{2}{\exp\left(\frac{\epsilon}{M(M-1)}\right) + 1}\right) .$$

Remark 8. *The factor $1/D(\epsilon)^2$ in the sample complexity bound of* MALLOWSKLD *grows fast with M. Therefore this algorithm is practical only for small $M (< 10)$. It is an interesting open question whether the KLD problem can be solved in a more efficient way for Mallows.*

7. Experiments

The experimental studies presented in this section are mainly aimed at showing advantages of our approach in situations where its model assumptions are indeed valid. To this end, we work with synthetic data. Yet, experiments with real data are presented in the supplementary material.

Doignon et al. (2004) introduced an efficient technique for sampling from the Mallows distribution. Based on Theorem 2, however, one can readily calculate the pairwise marginals for given parameters ϕ and $\widetilde{\mathbf{r}}$. Therefore, sampling the pairwise comparisons for a particular pair of objects i and j is equivalent to sampling a Bernoulli distribution with parameter $g(\widetilde{r}_i, \widetilde{r}_j, \phi)$.

7.1. The most preferred item (MPI)

We compared our MALLOWSMPI algorithm with other preference-based algorithms applicable in our setting, namely INTERLEAVED FILTER (IF) introduced by Yue et al. (2012) and BEAT THE MEAN (BTM) by Yue & Joachims (2011)[5]. While both algorithms follow a successive elimination strategy and discard items one by one, they differ with regard to the sampling strategy they follow.

Since the time horizon must be given in advance for IF, we run it with $T \in \{100, 1000, 5000, 10000\}$, subsequently referred to as IF(T). The BTM algorithm can be accommodated into our setup as is (see Algorithm 3 in (Yue & Joachims, 2011)).

We compared the algorithms in terms of their empirical sample complexity (the number of pairwise comparison until termination). In each experiment, the center ranking of the Mallows model was selected uniformly at random (since Mallows is invariant with respect to the center ranking, the complexity of the task is always the same). Moreover, we varied the parameter ϕ between 0.05 and 0.8. In Figure 2, the sample complexity of the algorithms is plotted against the parameter ϕ. As expected, the higher the value of ϕ, the more difficult the task. As can be seen from the plot, the complexity of MALLOWSMPI is an order of magnitude smaller than for the other methods. The empirical accuracy (defined to be 1 in a single run if the most preferred object was found, and 0 otherwise) was significantly bigger than $1 - \delta$ throughout.

The above experiment was conducted with $M = 10$ items. However, quite similar results are obtained for other values of M. The corresponding plots are shown in the supplementary material (see Appendix E).

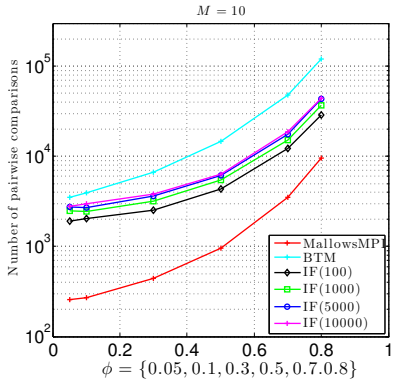


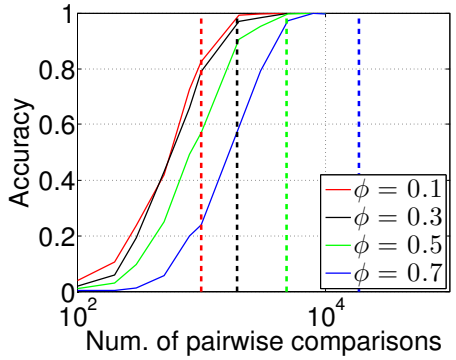
Figure 2. The sample complexity for $M = 10$, $\delta = 0.05$ and different values of the parameter ϕ. The results are averaged over 100 repetition.

7.2. The most probable ranking (MPR)

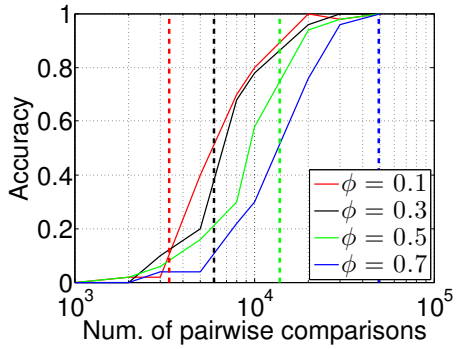
Cheng et al. (2009) introduced a parameter estimation method for the Mallows model based on the maximum likelihood (ML) principle. Since this method can handle incomplete rankings, it is also able to deal with pairwise comparisons as a special case. Therefore, we decided to use this method as a baseline.

We generated datasets of various size, consisting of only pairwise comparisons produced by a Mallows model. More specifically, we first generated random rankings according to Mallows (with fixed ϕ and center ranking selected uniformly at random) and then took the order of the two items

[5]The most naive solution would be to run the SUCCESSIVEELIMINATION algorithm (Even-Dar et al., 2002) with $Y_{i,1}, \ldots, Y_{i,M}$ as arms for some randomly selected i, where $Y_{i,j} = \mathbb{I}\{r_i < r_j$, where $\mathbf{r} \sim P(.|\phi, \widetilde{r})\}$. The problem with this approach is that by selecting i such that $\widetilde{r}_i = M$, the gap between the mean of the best and second best arm is very small (namely $p_{M,1} - p_{M,2} \leq (2(M-1)\phi^{M-1})/(1+\phi)$ based on Corollary 3). Therefore, the sample complexity of SUCCESIVEELIMINATION becomes huge.



(a) $M = 10$.



(b) $M = 20$.

Figure 3. The accuracy of the ML estimator versus the number of pairwise comparisons for various parameters ϕ. The horizontal dashed lines show the empirical sample complexity of MALLOWSMPR for $\delta = 0.05$. The results are averaged over 100 repetitions.

that were selected uniformly from $[M]$. We defined the accuracy of an estimate to be 1 if the center ranking was found, and 0 otherwise.

The solid lines in Figure 3 plot the accuracy against the sample size (namely the number n of pairwise comparisons) for different values $\phi \in \{0.1, 0.3, 0.5, 0.7\}$. We also run our MALLOWSMPR algorithm and determined the number of pairwise comparisons it takes until it terminates. The horizontal dashed lines in Figure 3 show the empirical sample complexity achieved by MALLOWSMPR for various ϕ. In accordance with Theorem 5, the accuracy of MALLOWSMPR was always significantly higher than $1-\delta$ (close to 1).

As can be seen, MALLOWSMPR outperforms the ML estimator for smaller ϕ, in the sense of achieving the required accuracy of $1-\delta$, whereas the accuracy of ML is still below $1-\delta$ for the same sample complexity. Only for larger ϕ, the ML approach does not need as many pairwise comparisons as MALLOWSMPR to achieve an accuracy higher than $1-\delta$. For $M = 20$, the advantage of MALLOWSMPR is even more pronounced (see Figure 3(b)).

8. Conclusion and future work

The framework of rank elicitation introduced and analyzed in this paper differs from existing ones in several respects. In particular, sample information is provided in the form of pairwise preferences (instead of individual evaluations), an assumption that is motivated by practical applications. Moreover, we assume a data generating process in the form of a probability distribution on total orders. This assumptions has (at least) two advantages. First, since there is a well-defined "ground truth", it suggests clear targets to be estimated and learning problems to be tackled, like those considered in this paper (MPI, MPR, KLD). Second, exploiting the properties of models such as Mallows, it is possible to devise algorithms that are more efficient than general purpose solutions.

Of course, this last point requires the model assumptions to hold in practice, at least approximately. This is similar to methods in parametric statistics, which are more efficient than non-parametric methods provided their assumptions are valid. An important topic of future work, therefore, is to devise a (Kolmogorov-Smirnov type) hypothesis test for deciding, based on data in the form of pairwise comparisons, whether the underlying distribution could indeed be Mallows. Although this is a challenging problem, it is arguably simpler than testing the validity of strong stochastic transitivity and stochastic triangle inequality as required by methods such as IF and BTM.

Apart from that, there is a number of interesting variants of our setup. First, ranking models other than Mallows can be used, notably the Plackett-Luce model (Plackett, 1975; Luce, 1959), which has already been used for other machine learning problems, too (Cheng et al., 2010; Guiver & Snelson, 2009); since this model is less restrictive than Mallows, sampling algorithms and complexity analysis will probably become more difficult. Second, going beyond pairwise comparisons, one may envision a setting in which the learner is allowed to query arbitrary subsets of items (perhaps at a size-dependent cost) and receive the top-ranked item as feedback.

Acknowledgments

This work was supported by the German Research Foundation (DFG) as part of the Priority Programme 1527, and by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).

References

Braverman, M. and Mossel, E. Noisy sorting without resampling. In *Proceedings of the nineteenth annual ACM-*

SIAM Symposium on Discrete algorithms, pp. 268–276, 2008.

Bubeck, S., Munos, R., and Stoltz, G. Pure exploration in multi-armed bandits problems. In *Proceedings of the 20th ALT*, ALT'09, pp. 23–37, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 3-642-04413-1, 978-3-642-04413-7.

Bubeck, S., Wang, T., and Viswanathan, N. Multiple identifications in multi-armed bandits. In *Proceedings of The 30th ICML*, pp. 258–265, 2013.

Busa-Fekete, R., Szörényi, B., Weng, P., Cheng, W., and Hüllermeier, E. Top-k selection based on adaptive sampling of noisy preferences. In *Proceedings of the 30th ICML, JMLR W&CP*, volume 28, 2013.

Cappé, O., Garivier, A., Maillard, O.-A., Munos, R., and Stoltz, G. Kullback-Leibler upper confidence bounds for optimal sequential allocation. *Submitted to the Annals of Statistics*, 2012.

Chen, X., Bennett, P. N, Collins-Thompson, K., and Horvitz, E. Pairwise ranking aggregation in a crowd-sourced setting. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 193–202, 2013.

Cheng, W., Hühn, J., and Hüllermeier, E. Decision tree and instance-based learning for label ranking. In *Proceedings of the 26th International Conference on Machine Learning*, pp. 161–168, 2009.

Cheng, W., Dembczynski, K., and Hüllermeier, E. Label ranking methods based on the plackett-luce model. In *27th ICML*, pp. 215–222, 2010.

Doignon, J., Pekeč, A., and Regenwetter, M. The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika*, 69(1): 33–54, 2004.

Even-Dar, E., Mannor, S., and Mansour, Y. PAC bounds for multi-armed bandit and markov decision processes. In *Proceedings of the 15th COLT*, pp. 255–270, 2002.

Fill, J. A. and Janson, S. Quicksort asymptotics. *Journal of Algorithms*, 44(1):4 – 28, 2002.

Flajolet, P. and Golin, M. J. Mellin transforms and asymptotics: The mergesort recurrence. *Acta Inf.*, 31(7):673–696, 1994.

Fligner, M. A. and Verducci, J. S. Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):359–369, 1986.

Gabillon, V., Ghavamzadeh, M., Lazaric, A., and Bubeck, S. Multi-bandit best arm identification. In *Advances in NIPS 24*, pp. 2222–2230, 2011.

Guiver, J. and Snelson, E. Bayesian inference for plackett-luce ranking models. In *Proceedings of the 26th ICML*, pp. 377–384, 2009.

Guo, S., Sanner, S., Graepel, T., and Buntine, W. Score-based bayesian skill learning. In *European Conference on Machine Learning*, pp. 1–16, September 2012.

Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

Jamieson, K.G. and Nowak, R.D. Active ranking using pairwise comparisons. In *Advances in Neural Information Processing Systems 24*, pp. 2240–2248, 2011.

Lu, T. and Boutilier, C. Learning mallows models with pairwise preferences. In *Proceedings of the 28th International Conference on Machine Learning*, pp. 145–152, 2011.

Luce, R. D. *Individual choice behavior: A theoretical analysis.* Wiley, 1959.

Mallows, C. Non-null ranking models. *Biometrika*, 44(1): 114–130, 1957.

Marden, John I. *Analyzing and Modeling Rank Data*. Chapman & Hall, 1995.

Plackett, R. The analysis of permutations. *Applied Statistics*, 24:193–202, 1975.

Urvoy, T., Clerot, F., Féraud, R., and Naamane, S. Generic exploration and k-armed voting bandits. In *Proceedings of the 30th ICML, JMLR W&CP*, volume 28, pp. 91–99, 2013.

Yue, Y., Broder, J., Kleinberg, R., and Joachims, T. The k-armed dueling bandits problem. *Journal of Computer and System Sciences*, 78(5):1538–1556, 2012.

Yue, Yisong and Joachims, Thorsten. Beat the mean bandit. In *Proceedings of the ICML*, pp. 241–248, 2011.